

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение высшего образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ»
(СПбГЭУ)

Факультет информатики и прикладной математики
Кафедра прикладной математики и экономико-математических методов

КУРСОВАЯ РАБОТА
по дисциплине:
«Численные методы»

Тема Расширение МДМ-метода на задачу мягкого отделения

Направление: 01.03.02 «Прикладная математика и информатика»

Направленность:
Прикладная математика и информатика в экономике и управлении

Обучающийся 2 курса группы ПМ-2101 очной формы обучения

Зелин Иван Алексеевич ______________
(подпись)

Проверил Соловьёва Н.А.

Должность к.ф.-м.н., доцент

Оценка ____________________ Дата 29 мая 2023

Подпись ____________________

Санкт-Петербург
2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Эта работа - расширение известного МДМ-метода для решения общей квадратичной задачи математической диагностики на задачу мягкого отделения.

1. Постановка задачи

Усеченной выпуклой оболочкой (RCH) n точек p_i с усекающим коэффициентом μ называется множество

$$R = \{\sum_{i=1}^{n} u_i * p_i \mid 0 \le u_i \le \mu, \sum_{i=1}^{n} u_i = 1\}. \tag{1.1}$$

На μ накладывается естественное ограничение

$$\frac{1}{n} \le \mu \le 1. \tag{1.2}$$

Когда $\mu = \frac{1}{n}$, RCH очевидно содержит единственный элемент - центр масс множества.

Пусть в пространстве $\mathbb{R}^n$ с евклидовой нормой есть векторы p_i и заданы множества $P_1 = \{p_i\}_{i=1}^{s}$ и $P_2 = \{p_i\}_{i=s+1}^{m}$. Пусть R_1 - RCH(P_1), R_2 - RCH(P_2) . Наша цель - найти вектор кратчайшего расстояния между R_1, R_2, то есть такие w_1, w_2, что

$$\|w_1 - w_2\| \to \min_{w_1 \in R_1, w_2 \in R_2,}. \tag{1.3}$$

При этом для любого решения w_1, w_2 разность $w = w_1 - w_2$ одинакова. Основой алгоритма является МДМ-метод[1].

2. Описание алгоритма

Пусть уже есть некоторое приближение w_1, w_2:

$$w_1 = \sum_{i=1}^{s} u_1[i] P_1[i] \in R_1, w_2 = \sum_{i=1}^{m-s} u_2[i] P_2[i] \in R_2. \tag{2.1}$$

Введём для обеих RCH вспомогательные множества:

$$M_k^{(1)} = \{j | u_k[j] > 0\}, M_k^{(2)} = \{j | u_k[j] < \mu\}. \tag{2.2}$$

Вычислим $c_1 = \{\langle P_1[i], w_1 - w_2\rangle\}_{i=1}^{s},\ c_2 = \{\langle P_2[i], w_2 - w_1\rangle\}_{i=1}^{m-s}$, пропорциональные проекциям вектора кратчайшего расстояния на элементы множеств.
Найдем для обоих наборов индексы $j_k^{min} \in M_k^{(1)}$ и $j_k^{max} \in M_k^{(2)}$ такие, что:

$$\begin{cases} \min\limits_i c_k[i] = c_k[j_k^{min}], \\ \max\limits_i c_k[i] = c_k[j_k^{max}]. \end{cases} \tag{2.3}$$

Пусть $\Delta_k = c_k[j_k^{max}] - c_k[j_k^{min}]$.
Если $\Delta_1 = \Delta_2 = 0$, данные приближения - решение (показано далее). Процесс окончен.
Иначе для большего из Δ_k "сместим" соответствующий ему w_k в направлении $p_k[j_k^{min}] - p_k[j_k^{max}]$ (то есть от элемента с наибольшей проекцией к элементу с наименьшей, передав тому часть коэффициента первого) на величину

$$q = \min\{\frac{\Delta_k}{\|p_k[j_k^{min}] - p_k[j_k^{max}]\|^2}, u_k[j_k^{max}], \mu - u_k[j_k^{min}]\}. \tag{2.4}$$

Оптимальность выбора q будет показана далее.
Коэффициенты $u_k[s]$ разложения w_k по $P_k[i]$ изменятся таким образом:

$$u_k^{new}[s] = \begin{cases} u_k[s] \text{ при } s \notin \{j_k^{min}, j_k^{max}\}, \\ u_k[s] + q \text{ при } s = j_k^{min}, \\ u_k[s] - q \text{ при } s = j_k^{max}. \end{cases} \tag{2.5}$$

Выполним следующую итерацию с обновленным w_k.

3. Формализация

Запишем двойственную задачу[1] к (1.3).

$$\begin{cases} \frac{1}{2}\langle A^T A u, u\rangle - \sum\limits_{i=1}^{m} u_i \to \min, \\ \sum\limits_{i=1}^{m} \xi_i u_i = 0, \\ 0 \le u_i \le \mu, i \in 1:m, \end{cases} \tag{3.1}$$

где $\xi_i = 1$ при $i \in 1:s$ и $\xi_i = -1$ при $i \in s+1:m$,

$u_{1:s}, u_{s+1:m}$ - коэффициенты разложения приближения w_1, w_2 по векторам из $p_{1:s}$ и $p_{s+1:m}$ соответственно,

A - матрица со столбцами $\xi_1 p_1 ... \xi_m p_m$.

Покажем, что действительно $\Delta_1 = \Delta_2 = 0 \Leftrightarrow$ план w_1, w_2 - оптимален.

Доказательство. Сформулируем для задачи условие Каруша-Куна-Таккера[1]:

$$\begin{cases} A^T A u - e = \alpha\xi + \sum\limits_{i=1}^{m} \beta_i e_i - \sum\limits_{i=1}^{m} \gamma_i e_i, \\ \beta_j u_j = 0, \beta_j \ge 0,\ j \in 1:m, \\ \gamma_j(\mu - u_j) = 0, \gamma_j \ge 0,\ j \in 1:m, \end{cases}$$

α, β, γ – наборы двойственных переменных, e – единичный вектор длины m (3.2)

Аналогично представленным в предыдущей части (2.2) $M_k^{(1)},\ M_k^{(2)}$ введём вспомогательные множества

[1] Подробно о двойственной задаче и условии ККТ: www.cs.cmu.edu/~ggordon/10725-F12/slides/16-kkt.pdf .

$$
\begin{gathered}
I_+^{(1)} = \{j \in 1:s | u_j = 0\}, \quad I_+^{(2)} = \{j \in 1:s | 0 < u_j < \mu\}, \\
I_+^{(3)} = \{j \in 1:s | u_j = \mu\}, \\
I_-^{(1)} = \{j \in s+1:m | u_j = 0\}, \quad I_-^{(2)} = \{j \in s+1:m | 0 < u_j < \mu\}, \\
I_-^{(3)} = \{j \in s+1:m | u_j = \mu\}.
\end{gathered} \tag{3.3}
$$

Обозначим $w = w_1 - w_2 = Au = \sum_{i=1}^{m} \xi_i u_i p_i$. Тогда $A^T Au[i] = \langle p_i, w \rangle$. Условие (3.2) примет вид:

$$
\langle p_i, w \rangle = \begin{cases} 1 + \alpha\xi_i + \beta_i \text{ при } i \in I_+^{(1)} \cup I_-^{(1)}, \\ 1 + \alpha\xi_i \text{ при } i \in I_+^{(2)} \cup I_-^{(2)}, \\ 1 + \alpha\xi_i - \gamma_i \text{ при } i \in I_+^{(3)} \cup I_-^{(3)}. \end{cases} \tag{3.4}
$$

В текущих обозначениях использовавшиеся в (2) Δ_1 и Δ_2 выглядят следующим образом:

$$
\Delta_1 = \max_{I_+^{(2)} \cup I_+^{(3)}} \langle p_i, w \rangle - \min_{I_+^{(1)} \cup I_+^{(2)}} \langle p_i, w \rangle, \tag{3.5}
$$

$$
\Delta_2 = \max_{I_-^{(2)} \cup I_-^{(3)}} \langle p_i, w \rangle - \min_{I_-^{(1)} \cup I_-^{(2)}} \langle p_i, w \rangle. \tag{3.6}
$$

Теперь очевидно, что $\Delta_{max} = 0 \Leftrightarrow$ выполнено условие (3.2). □

Покажем также оптимальность выбора смещения q (2.4) на заданной итерации с приближением w.

Доказательство. Пусть $\Delta_1 > \Delta_2$. Обозначим

$$
w(q) = w - q(p_{j_1^{max}} - p_{j_1^{min}}), \quad q \in [0, u_{j_1^{max}}]. \tag{3.7}
$$

Тогда

$$\|w(q)\|^2 = \|w_k\|^2 - 2q\Delta_1 + q^2\|p_{j_1^{max}} - p_{j_1^{min}}\|^2 \tag{3.8}$$

Абсолютный минимум $\|w(q)\|^2$ достигается при

$$q = \frac{\Delta_1}{\|p_{j_1^{max}} - p_{j_1^{min}}\|^2}. \tag{3.9}$$

Из определения RCH (1.1) и описанием смещения (2.5) очевидны неравенства

$$u_{j_1^{min}} - \mu \le q \le u_{j_1^{max}}. \tag{3.10}$$

Поэтому

$$q = \min\{\frac{\Delta_1}{\|p_{j_1^{max}} - p_{j_1^{min}}\|^2}, u_{j_1^{max}}, \mu - u_{j_1^{min}}\}. \tag{3.11}$$

Случай $\Delta_2 > \Delta_1$ рассматривается аналогично. □

4. Сходимость алгоритма

Наш алгоритм является обобщением МДМ-метода для решения общей квадратичной задачи математической диагностики ($\mu = 1$), чья сходимость доказана в [2]. Сходимость представленного алгоритма проверена эвристически.

5. Связь с мягким SVM-отделением

Задача мягкого SVM-отделения ставится следующим образом [4]:

$$\begin{gathered} \frac{1}{2}\|w\|^2 + \mu \sum_{j=1}^{m} \eta_j \Rightarrow \text{min}, \\ \xi_j(\langle w, p_j \rangle + \beta) \ge 1 - \eta_j, \quad j \in 1:m, \\ \eta_j \ge 0, \quad j \in 1:m. \end{gathered} \tag{5.1}$$

Здесь η_j характеризует величину ошибки на p_j. Двойственная к этой задаче совпадает с (3.1) [5].

Пусть $\exists\, w_1, w_2$ решение задачи (1.3). Пусть $w = w_1 - w_2$. По теоремам двойственности у задачи (5.1) есть решение w^*, β^*, причём w - компонента w^* этого решения.

Из условий (5.1)

$$\begin{gathered} \beta^* \geq -\langle w^*, p_i \rangle, \quad i \in 1:s, \\ \beta^* \leq -\langle w^*, p_i \rangle, \quad i \in (s+1):m, \end{gathered} \tag{5.2}$$

откуда

$$\max_{i \in 1:s} \{1 - \langle w^*, p_i \rangle\} \leq \beta^* \leq \min_{i \in (s+1):m} \{-1 - \langle w^*, p_i \rangle\}. \tag{5.3}$$

Таким образом представленный алгоритм может применяться для решения задачи нестрогого отделения.

6. Практическое применение

Алгоритм позволяет решать широкий спектр задач и в частности имеет большое применение в машинном обучении. В задаче классификации является аналогом алгоритмов SVM-light и SMO.

Рассмотрим задачу бинарной классификации. Пусть есть множество объектов $\mathbb{G}$, каждый из которых относится к одному из двух классов, но информация об отношении объектов к классу есть лишь для $\mathbb{H} \subset \mathbb{G}$. Есть также признаковое пространство $\mathbb{F}$ и отнощения $\mathbb{A} : \mathbb{G} \to \mathbb{F},\ \mathbb{B} : \mathbb{H} \to \{0,1\}$. Пользуясь этим, требуется восстановить $\mathbb{C} : \mathbb{G} \to \{0,1\}$.

Если множества соответствующих объектам признаков линейно разделимы, задачу можно решить, применив МДМ, SVM-light, SMO методы.

Классический способ решения нелинейной задачи такими алгоритмами - преобразовать признаки из $\mathbb{F}$ таким образом, чтобы линейное разделение всё же было. Часто это приводит к увеличению размерности $\mathbb{F}$.

Описанный в работе метод позволяет разделять линейно неразделимые множества (рисунок 1) с сохранением размерности и удалением из рассмотрения

наиболее вырожденных элементов (рисунок 2). Мерой вырожденности можно считать удалённость элемента от центра масс множества.

Помимо изложенного, в рамках работы реализованы классификатор множеств в $\mathbb{R}^n$ на основе описанного метода и визуализатор результатов классификации (рисунок 3), (рисунок 4).

Список литературы

[1] B. F. Mitchell, V. F. Dem'yanov, and V. N. Malozemov: *Finding the point of a polyhedron closest to the origin*

[2] Малоземов В. Н., Соловьева Н. А.: *МДМ-метод для решения общей квадратичной задачи математической диагностики*

[3] Ming Zeng, Yu Yang, Junsheng Cheng: *A generalized Mitchell-Dem'yanov-Malozemov algorithm for one-class support vector machine*

[4] Воронцов К. В.: *Лекции по методу опорных векторов*

[5] Малоземов В. Н., Плоткин А. В.: *SVM-метод: мягкое отделение*

ПРИЛОЖЕНИЯ

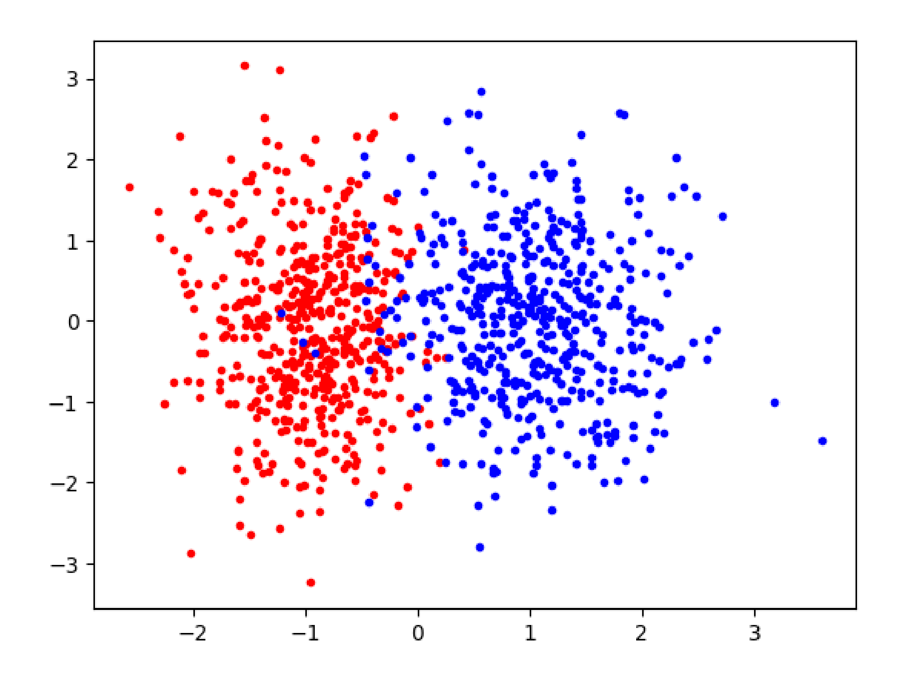


Рисунок 1 – Линейно неразделимые множества

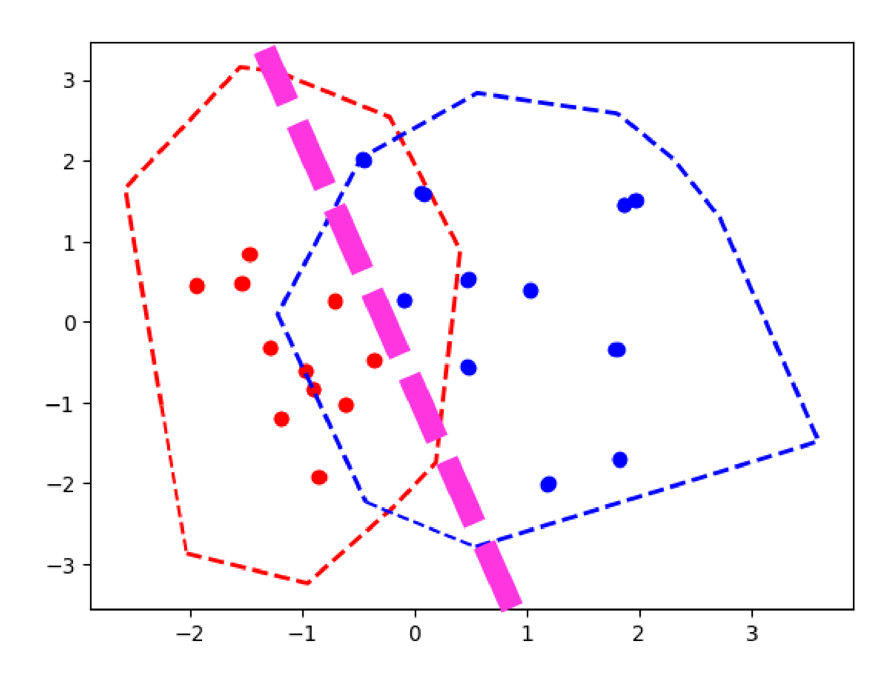


Рисунок 2 – Усечение: появилась возможность линейного разделения

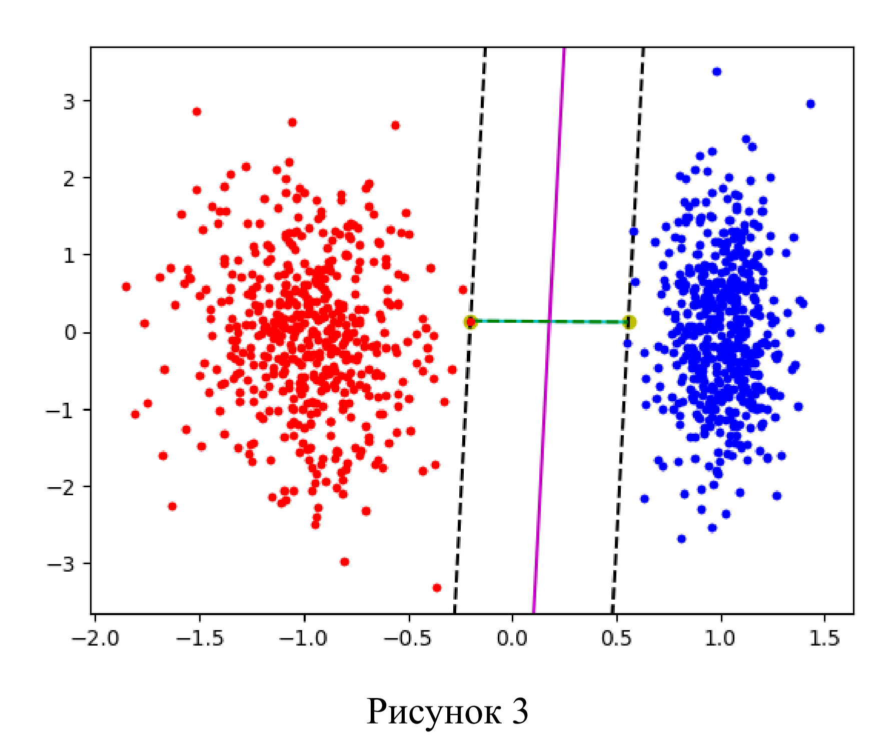


Рисунок 3

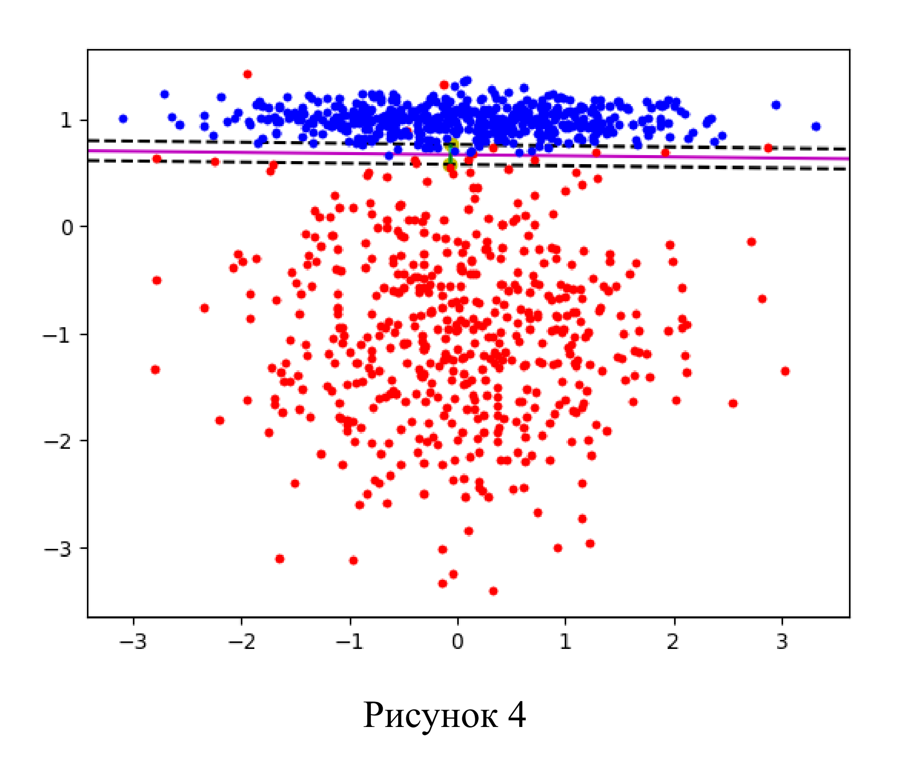


Рисунок 4